ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 31 (2014) 281 – 290

Procedia
Computer Science

2nd International Conference on Information Technology and Quantitative Management, ITQM 2014

A Method Based on One-Class SVM for News Recommendation

Limeng Cui[a,b,*], Yong Shi[a,b]

[a]University of Chinese Academy of Sciences, Beijing, 100049, China
[b]Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, 100190, China

Abstract

In order to provide intelligent recommendation and personalized service for users on news website, this paper presents a method based on One-Class SVM for news recommendation algorithm. By analyzing the news webpages and user's browsing history, and by building One-Class SVM model, this algorithm can recommend news for user. The main work of this paper is to study this news recommendation algorithm and to show its experimental results under Dot NET platform. First, this algorithm preprocesses the webpages from Sogou Labs, each of which has its inherent domain and builds One-Class SVM models for these domains. Next, it builds user interest models for each user by analyzing their browsing histories. Then it finds the user's most interested domains by comparing each domain models and user interest model. Finally, it utilizes the webpages of these domains and user's browsing history to build One-Class SVM model to calculate the most relevant webpages to user interest, and recommends these webpages to user. This algorithm takes the lead in calculate the similarity between user interests and webpages using One-Class SVM model and apply hierarchical model to make the results more accurate. From the results, we can find that this algorithm is running pretty well.

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.

Keywords: News recommendation; One-Class SVM; Hierarchical recommendation algorithm; Similarity calculation; Vector Space Model (VSM)

1. Introduction

Recently, the Internet is becoming the main route of spreading news due to the growing information technology. Billions of information would be produced globally. According to the 31th China Internet Network Development

* Corresponding author. Tel.: +86-13811532932.
E-mail address: lmcui932@163.com (Limeng Cui), yshi@ucas.ac.cn (Yong Shi).

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.
doi:10.1016/j.procs.2014.05.270

Statistic Report, as of December 2012, Chinese webpage number is 122.7 billion, grew by 41.7% with the same period in 2011. This fully shows that the growth of the Internet information is exploding. In the face of such a tremendous amount of information, how to discover useful information quickly and precisely is the crucial problem to be solved currently.

News websites are emerging constantly, which although provide users with giant amounts of information and news, but also make users faced with the "information overload" problem. While searching for the information needed, users are forced to browse massive amount of irrelevant webpages.

It is hard for users to dig out useful and high-quality information by their own. Search engine technology finds related document based on the keywords users input. But studies have found that this method returns a considerable amount of documents, but one of only a fraction is related to the user's query[1]. News recommendation technology is practiced to solve "information overload" problem. Content-based recommendation system plays an important role in recommendation systems, and is suitable for user personalized recommendation[2].

This paper presents a method based on One-Class SVM for news recommendation algorithm. This method, by introducing One-Class SVM method into the calculation of the similarity between webpage and user, can recommend news for users by using users' browsing history.

Firstly, we analyze and summarize the basic ways of recommendation system. According to domestic and international related literature, we have summarized the methods of recommendation system and personalized recommendation, such as content-based filtering and collaborative filtering.

In the next place, aiming at the shortcomings of the traditional single-layer VSM news recommendation algorithm, we apply hierarchical algorithm to it. Traditional single-layer algorithms only calculate the similarity between webpage and user, and recommend news for users according to the list ranked in order of the biggest similarities. In addition, this model only calculates the similarity between news in the domains (the category of news) users are interested in and user interest models, not all the news. So we can consider this model is time-saving. Hierarchical model can make the results more accurate.

Besides, we use One-Class SVM method to calculate the similarity between webpage and user interest model. The experimental result is good. And more important is that there has been no study on this method is applied to news recommendation.

Finally, a news recommendation system has been designed and implemented. We have completed each module and conduct the performance evaluation.

2. Related Works

Broadly, existing recommendation system technologies can be divided into two categories: content-based filtering and collaborative filtering.

Based on a profile of user interests and preferences, content-based recommendation system can recommend items that may be of interest or value to the user[2], which has also been widely used in the aspects of information retrieval and information filtering. Personal WebWatcher is a personalized news recommendation system which was introduced by Carnegie Mellon University. This system can determine user interest by his browsing history. It infers that browsed records are interesting, contrarily vapidity.

Collaborative filtering recommendation system generally adopts the adjacent technology, using the historical record of users to measure the distance between them, thus to recommend for target user based on his neighbor's preferences.

The content-based filtering and collaborative filtering each has their merits and limitations. Some research tried to combine these two methods and achieved encouraging results such as WebSIFT[3] and Anatagonomy[4].

For news websites, users are relatively independent. Due to the absence of the concept "neighbor", collaborative filtering algorithm lost its advantages.

The mainstream news recommendation technology is based on Vector Space Model (VSM). Many scholars have made improvement on the classic recommendation algorithm and obtained good results. Here are several news recommendation algorithms.

Single-layer news recommendation algorithm based on VSM expresses the user interest and webpage respectively by a spatial vector comprised of a characteristic word and weight. Weight calculation is by TF-IDF.

Then, through calculation of the cosine similarity of the two vectors, the similarity between webpage and user interest is expressed. And highly similar articles are recommended to the users.

Literature[8] proposes a hierarchical news recommendation algorithm based on VSM that recommends webpages one by one of each domain in the descending order of the similarity between domain and user interest. This algorithm ensures more accurate recommended webpages. In addition, it will save much time if the webpages in domains that are highly similar to the user interest are only recommended.

In order to improve the drawbacks of the algorithm mentioned above, literature[9] puts forward an improved hierarchical new recommendation algorithm based on VSM. It also firstly calculates the similarity between domain and user interest and then, in the descending order of the foresaid similarity, the similarity between each webpage and the user is calculated. But what is different is that it proposes to use independent similarity to replace that of webpage and user interest when recommending webpages to the user. Independent similarity is the weighted sum of the similarity between domain and user interest, as well as of the webpage and user interest. Thus the domain, webpage and user interests are comprehensively considered. Recommendation performance is improved. And accuracy of recommendation is enhanced in the case of many categories of webpages.

Most existing algorithms are single-layer recommendation algorithms based on VSM. Specifically, each webpage and user interest is expressed as a vector of a characteristic word and weight. The weight is calculated by TD-IDF. Finally, webpages with high similarity with the user model are recommended through calculation of the cosine similarity of the vectors. Two disadvantages of this algorithm: firstly, in the case of large-scale webpages, the calculation amount of similarity of webpages and user interest one by one as well as sorting will be enormous. In addition, it's unnecessary to recommend other unrelated webpages to users who have dedicated interests.

Many recommendation systems adopt hierarchical algorithms. Algorithms like SVM only solve double class issues and consider webpages unread by users are not interested to them and vice versa. This assumption is incorrect, as it does not show that the user is not interested in the webpages that he hasn't read. Therefore, single class algorithm, like One-Class SVM, is highly applicable in news recommendation. Accuracy is largely increased.

The algorithm we proposed has made improvements based on the above two drawbacks.

3. A Method Based on One-Class SVM for News Recommendation

To make up for large amount of calculation of the traditional single-layer algorithm, we use the hierarchy algorithm. The algorithm we proposed firstly calculates the similarity between domain and user interest. Select the domains which have the biggest similarities as the domains that user is interested in. Then, calculate the similarity between webpage in these domain and user interest. Finally, recommend webpages to user in the descending order of the foresaid similarity. We introduce One-Class SVM to the calculation of the similarity of domain and user interest, and the similarity between webpage and user interest.

3.1. One-Class SVM for News Recommendation

SVM can create a non-linear decision boundary by projecting the data through a non-linear function $\phi(\cdot)$ to a space with a higher dimension. This means that data points which can't be separated by a straight line in their original space R are "lifted" to a feature space F where there can be a "straight" hyperplane that separates the data points of one class from another. This method can be understood as Fig. 1.

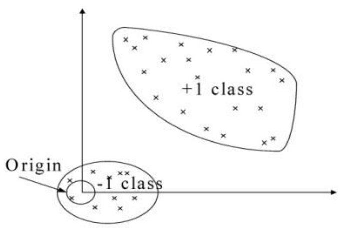

Fig. 1. One-Class SVM Model

Suppose that we have a set of N-dimensional data points $x_j \in R^N (j = 1,2,...,l)$. And assume that we have a mapping $\phi(\cdot): R^N \mapsto F$ which maps the data points into a higher-dimensional feature space, denoted by F. For simplicity, we denote the mapped image $\phi(x_j)$ as ϕ_j, hereinafter. Let $w \in F$ and $\rho \in R$. The purpose of the 1-SVM is to calculate a hyperplane that holds most of the data points in its positive side[5], i.e., $\langle w, \phi_j \rangle - \rho > 0$. We can find the hyperplane by maximizing the minimum Euclidean distance $\frac{\rho}{\|w\|}$ between the origin and the target data points.

Because the data set will inevitably contain abnormal points and noise points, in order to maintain the accuracy of the algorithm, slack variable will usually be introduced to reduce the error brought by the abnormal points and noise points.

Based on above, One-Class SVM is to conduct the following optimization problem:

$$\min \frac{1}{2} \| w \|^2 + \frac{1}{vl} \sum_{j=1}^{l} \xi_j - \rho \tag{1}$$

$$s.t. \langle w, \phi_j \rangle \geq \rho - \xi_j, \xi_j \geq 0, j = 1,2,...,l$$

where $x_j \in R^r$ is the keyword; $w \in R^r$ is the normal vector of the hyperplane, which is also the weights of the keyword; $\rho \in R$ is offset value; $v \in (0,1]$ is a predefined positive parameter which need to be set manually[10], l stands for the number of the points in dataset, i.e., the number of webpages. Let (w^*, ρ^*) denote an optional solution of (1).

In particular, $\frac{1}{vl} \sum_{j=1}^{l} \xi_j$ is a slack function, which can eliminate the harm of the abnormal points.

When a data point belongs to the negative side of the hyperplane, i.e., $\langle w^*, \phi_j \rangle + \rho^* < 0$, its pattern can be considered different from the given single class of data points[5].

Suppose that we now have a webpages set $P = \{p_1, p_2, ..., p_l\}$, for each webpage $p_j \in P$, the associated feature vectors $\phi_j \in F$ are obtained. Besides, $P(a) \subseteq P$ is a subset of indices that are rated as webpages that user may be interested in, or that have actually been read by this user. We assume that $P(a)$ consists of l' webpages and is defined as $P(a) = \{1, 2, ..., l'\}$, which is treated as a set of the single class of data points in the problem. Let (w^*, ρ^*) denote an optional solution of (1). Then, for each webpage p_i that has not been read, i.e., $p_i \in P - P(a)$.

The distance from the hyperplane calculated as $(w^* \cdot \phi_i + \rho^*) / (w^*, w^*)$ can be used as an interest score of the webpage p_i. Ignoring the constants, one can use the real number $w^* \cdot \phi_i$ as a score to rank the webpage p_i for a specific user u_x.

We need a function $K(\cdot,\cdot)$, or a mapping $\phi(\cdot)$ as the kernel function, to project data that is not linearly separable into a higher dimensional space can make it linearly separable. See Fig. 2.

Since the outcome of the decision function only relies on the dot-product of the vectors in the feature space F, it is not necessary to perform an explicit projection to that space. As long as a function $\phi(\cdot)$ has the same results, it can be used instead. Popular choices for the kernel function are linear, polynomial, sigmoidal but mostly the Gaussian Radial Base Function (RBF):

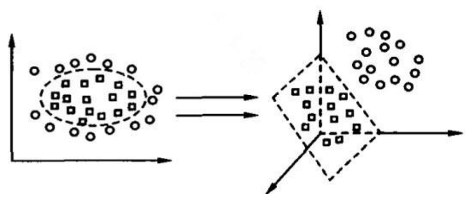

Fig. 2. Projecting

$$K(X_i, X_j) = exp(-\gamma \left\| X_i \cdot X_j \right\|)^2 \tag{2}$$

where γ is a kernel parameter and $\left\| X_i \cdot X_j \right\|$ is the dissimilarity measure.

RBF kernel is suitable for high or low dimension, large or small sample and so on. It has a wide domain of convergence. We assume that these features of RBF kernel make it the ideal function in this case. So in this algorithm, we use the RBF kernel. Since the most important parameter in RBF kernel is γ, we use 5-fold cross validation to find it.

3.2. A Hierarchy Method Based on One-Class SVM for News Recommendation

The algorithm proposed can be divided into four steps:

- Construct One-Class SVM model for user u and k domains.
- Calculate the similarity between user model and each domain models.
- Select m domains which have the biggest similarities, calculate the similarity between user model and the webpages in them.
- Select the webpages which have the biggest similarities, and recommend them to user.

3.2.1. Construct One-Class SVM model for user

For user u, construct One-Class SVM user interest model by using webpages set $Q = \{q_1, q_2, ..., q_z\}$ that have been read. Conduct the following optimization problem:

$$\min \frac{1}{2}\|w_u\|^2 + \frac{1}{vz}\sum_{i=1}^{z}\xi_i - \rho \tag{3}$$

$$s.t. \langle w_u, \phi_{q_i} \rangle \ge \rho - \xi_i, \xi_i \ge 0, i = 1, 2, ..., z$$

Find w_u as the feature vector to represent user interest.

3.2.2. Construct One-Class SVM model for domain

The webpages in the i^{th} domain can be represented as $P^{(i)} = \{p_1^{(i)}, p_2^{(i)}, ..., p_{l_i}^{(i)}\}$. And $l_1 + l_2 + ... + l_k = n$. n is the total number of webpages we have. Assume we have k domains. The domains set can be represented as $D = \{d_1, d_2, ..., d_k\}$. For each domain, construct One-Class SVM domain model. Conduct the following optimization problem:

$$\min \frac{1}{2}\left\|w_{d_i}\right\|^2 + \frac{1}{vl_i}\sum_{j=1}^{l_i}\xi_j - \rho \tag{4}$$

$$s.t. \langle w_{d_i}, \phi_{p_{l_i}^{(i)}} \rangle \ge \rho - \xi_j, \xi_j \ge 0, j = 1, 2, ..., z$$

Find w_d as the feature vector to represent domain.

3.2.3. Calculate the similarity between user interest and domain

The similarity between domain D_i and user u can be calculated as:

$$sim(d_i, u) = \frac{w_{d_i} \cdot w_u}{\left\|w_{d_i}\right\| \left\|w_u\right\|} \tag{5}$$

3.2.4. Calculate the similarity between user interest and webpage

Sort the similarities between domain and user interest in descending order. Select the top m domains which most resemble user interest. Sort the similarities between the webpages in these domain and user interest. The similarity can be calculated as:

$$sim(p_i, u) = \langle w_u, \phi_{p_i} \rangle - p_u \quad (6)$$

Then sort these webpages in the descending order of similarity. Select the top M webpages and recommend them to user.

3.3. Procedure

We combine the method above with the hierarchy thought and propose a hierarchy method based on One-Class SVM for news recommendation. This method first conduct One-Class SVM model for each domain and user interest, and find the feature vectors for each domain and user interest. Next, calculate the similarity between each domain and user interest and sort these domains in the descending order of similarity. Then, select the top m domains and calculate the similarity between the webpages in these domains and user interest. Finally, select the top M webpages and recommend them for user.

The algorithm of this method is summarized in Algorithm 1.

Algorithm 1: The algorithm of a hierarchy method based on One-Class SVM for news recommendation

Input: Domains set $D = \{d_1, d_2, ..., d_k\}$, webpages set $P^{(i)} = \{p_1^{(i)}, p_2^{(i)}, ..., p_{l_k}^{(i)}\}$ for each domain, webpages set $Q = \{q_1, q_2, ..., q_z\}$ which user u have read and parameters v, m, M.

Output: Webpages set S that will be recommended for user u.

Begin Set $DomainSimSet, PageSimSetS = \phi$;

Construct One-Class SVM model for user u base on parameter Q, v;

For all $d \in D$ do

Construct One-Class SVM model for each domain base on parameter $P^{(d)}, v$;

  Calculate the similarity $sim(d, u)$ between each domain and user interest base on w_u, w_d;

  $DomainSimSet \leftarrow DomainSimSet \cup (d, sim(d, u))$

End for

$Sort(DomainSimSet)$;

$D' = SelectTop(DomainSimSet, m)$;

For all $d \in D'$ do

  For all $p \in P^{(d)}$ do

    Calculate $sim(p, u)$;

    $PageSimSet \leftarrow PageSimSet \cup (p, sim(p, u))$;

  End for

End for

$Sort(PageSimSet)$;

$S = SelectTop(PageSimSet, M)$;

End

4. Experiments

4.1. Experimental Data

In this section, we conduct experimental studies of the algorithm we proposed. The data comes from the whole network data of 2012 issued by Sogou Labs, covering domestic news data from international, PE, social and entertainment channels and the like of several news portals from June to July 2012.

We randomly selected 5 registered users from the network and obtained pages that they browsed in the latest month through records. And then, for each user, the researcher randomly extracted 200 pages visited by them and took 100 pages as training data while the others were used as test data. After word processing of all pages extracted, every web page was represented in vector format finally.

We compare the similarities between the web pages gained by different recommendation algorithms and the actual pages visited by users so as to measure the pros and cons of such algorithms. For example, if certain algorithm calculates the training set of pages and gets some recommendation results little different from the test set, it indicates that this algorithm has outstanding recommendation effects.

The values m and M can be defined manually. In this paper, we set $m=3$ and $M=10$. It will be very time-consuming when m and M grow bigger, while the precision doesn't seem to be improved.

4.2. Evaluating Indicators

- Precision: the number of web pages recommended correctly divide by the total pages recommended.
- Recall: the number of web pages recommended correctly divide by the number of pages that users visited.
- F1 Score: a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score

Among them, web pages recommended correctly is defined as the intersection set of the pages recommended and the actual pages that users visited, while the latter, pages visited, is defined as the web pages from the test set.

4.3. Experimental Design

4.3.1. Comparison of Recommendation Algorithms

This experiment applied 5-fold cross validation for every user's data and randomly divided 200 web pages into 5 parts. The researcher took 4 parts thereof as training data with the rest part as test data and executed recommendation algorithm under the same parameter setting. We calculated the recommendation precision and recall according to the recommendation results of certain algorithm. And then, the researcher took the average of the 5 results and calculated final F1 score. For the selection of optimal parameter, we applied grid search and ran 5-fold cross validation mentioned above for each candidate parameter set. Finally, the candidate set with maximum F1 score was picked as the optimal parameter and the final experimental result.

In this experiment, we compare the hierarchy One-Class SVM recommendation method (HOCSVM) proposed in the article with directly used One-Class SVM (OCSVM), a traditional content-based recommendation method single-layer vector space (SLVSM), hierarchy vector space (HVSM) mentioned in literature[8], and hierarchy space using independent similarity (HVSMIS) mentioned in literature[9].

According to the foregoing experimental results, the algorithm proposed in this article is superior to other recommendation algorithms at precision, recall and F1 score from the viewpoint of users. It indicates that the method can establish the characteristics of reading habits for specific users by use of One-Class SVM to better fit the user's news browsing features and give news recommendations compliant with user's reading habits.

We summarized the foregoing recommendation results of each user and made sorting comparison under certain evaluation indicator so as to get the results of algorithms comparison as follows.

Fig. 3. shows the precision comparison under different algorithms and it is obvious from the figure that the web pages recommended by the algorithm herein are more likely to fit the favorite fields of customers. Therefore, the algorithm boasts better precision than others. Fig. 4. shows the recall comparison under different algorithms. And the web pages recommended by the algorithm herein better conform to the reading habits of users according to the

figure, because the algorithm establishes One-Class SVM models respectively for fields and customers. Therefore, it boasts better recall than the others discussed. Fig. 5. shows the F1 score under different algorithms.

Based on the foregoing analysis, the algorithm hereof shows better results at precision and recall so that it also has a better performance at F1 score.

4.3.2. Algorithm Parameter Adjustment

Through algorithm parameter adjustment, the researcher calculated the corresponding precision and recall of each recommendation algorithm on user1's data set.

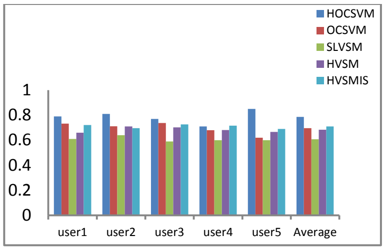

Fig. 3. Precision comparison under different algorithms

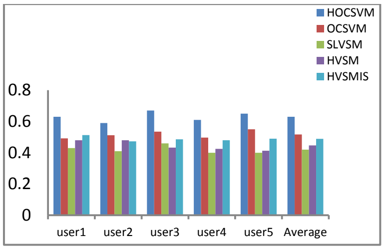

Fig. 4. Recall comparison under different algorithms

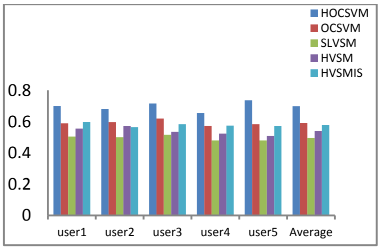

Fig. 5. F1 score comparison under different algorithms

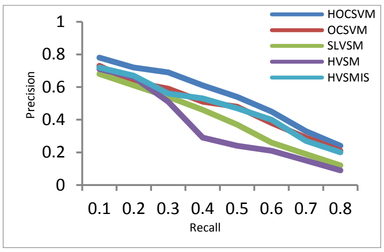

Fig. 6. Precision and Recall after parameter adjustment

According to the experimental results, which is shown in Fig. 6., the precision of single-layer method is lower than those of most algorithms under the same recall, but it is better than hierarchy algorithms when the precision is greater than 0.3. This may be led to by the fact that layered algorithms fail to effectively identify users' preferences under higher recalls so that corresponding precisions are lowered. The sole application of One-Class SVM recommendation method leads to direct recommendation without fields distinguished because of no consideration for users' favorite fields. Therefore, its precision is slightly lower than that of the method hereof. Moreover, the method hereof is superior to the others at precision under the same recall as it not only considers the users' favorite fields but also conforms to their reading habits.

5. Conclusion and Future Work

Considering the current status of news recommendation system, this paper proposes an algorithm based on One-Class SVM for news recommendation, which can recommend news information for users according to their browsing history, and the experiments show satisfactory results. The paper first presents several mainstream algorithms today, analyzes their limitations and studies the application of One-Class SVM in recommendation systems. Then, the concept of hierarchy recommendation is introduced into the proposed algorithm, i.e., the similarity between user interest and domain is calculated first, followed by the similarity between webpages in these domain and user interest, and then the recommendation is made accordingly. The paper is the first to propose the similarity calculation method by using One-Class SVM and apply hierarchy model to increase the accuracy of recommendation results. Following this, multiple experiments are conducted to test the effectiveness of the news recommendation algorithm based on One-Class SVM and the current mainstream algorithms. The results indicate that the method considerably outperforms the others.

Moreover, it is found that some topics in the recommendation algorithm are still worth further study:

- Currently, most of the recommendations on e-commerce websites are only for the specific structured items in the sites, such as movie and music. This shows that a more diversified information recommendation approach would be a meaningful research direction.
- This paper focuses on the comparison between One-Class-SVM-based Chinese news recommendation algorithm and other algorithms, but it has not included any researches on the setting of its parameters. This should be a follow-up for us in future.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grants (Grant No. 71110107026, 71331005) and the CAS/SAFEA International Partnership Program for Creative Research Teams.

References

1. Hu W C, Chen Y, Schmalz M S, et al. An overview of world wide web search technologies. *The proceedings of 5th World Multi Conference on Systems*, Cybernetics, Informatics, SCI2001, Orlando, Florida. 2001: 22-25.
2. Liu J, Dolan P, Pedersen E R. Personalized news recommendation based on click behavior. *Proceedings of the 15th international conference on Intelligent user interfaces.* ACM, 2010: 31-40.
3. Cooley R, Tan P N, Srivastava J. Websift: the web site information filter system. *Proceedings of the Web Usage Analysis and User Profiling Workshop.* 1999, 8.
4. Kamba T, Sakagami H, Koseki Y. ANATAGONOMY: a personalized newspaper on the World Wide Web. *International Journal of Human-Computer Studies*, 1997, 46(6): 789-803.
5. Yajima Y. One-class support vector machines for recommendation tasks. *Advances in Knowledge Discovery and Data Mining.* Springer Berlin Heidelberg, 2006: 230-239.
6. Lin S M, Wang G S, Chen Y Q. User modeling and feature selection in personalized recommending system. *Computer Engineering*, 2007, 33(17): 196-198.
7. Manevitz L M, Yousef M. One-class SVMs for document classification. *The Journal of machine Learning research*, 2002, 2: 139-154.

8. de Campos L M, Fernández-Luna J M, Gómez M, et al. A decision-based approach for recommending in hierarchical domains. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Springer Berlin Heidelberg, 2005: 123-135.
9. SHAO Hua, GAO Feng-Rong, XING Chun-Xiao, JIANG Li-Hua. A Hierarchical Webpage Recommendation Algorithm Based on Vector Space Model. *Computer Science*, 2006, 33(11): 86-88.
10. Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution. *Neural computation*, 2001, 13(7): 1443-1471.
11. Yong Shi, Yingjie Tian, Gang Kou, Yi Peng, Jianping Li. *Optimization Based Data Mining: Theory and Applications*. Springer, 2011.
12. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering*, IEEE Transactions on, 2005, 17(6): 734-749.
13. Lin S, Liu Z. Parameter selection in SVM with RBF kernel function. *Journal-Zhejiang University of Technology*, 2007, 35(2): 163.
14. Wang X M, Peng H. One-class SVM Based on Data Distribution. *Journal of Xihua University* (Natural Science Edition), 2012, 1: 009.
15. FENG A, LIU X, SUN T. Embedding Target Data's Structural Distribution Information into One-Class SVM and Its Linear Programming Algorithm. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2009, 6: 002.